Grounding Language in Object-Centered Affordance

Mark Steedman, University of Edinburgh

http://www.inf.ed.ac.uk/~steedman

Humanoids Conference: Workshop 7 Paris Dec 2009

Outline

- Introduction
- I: Planning
- II: Grammar
- Conclusion

Introduction

- There is a long tradition associating language and other serial cognitive behavior with an underlying motor planning mechanism (Piaget 1936, Lashley 1951, Miller *et al.* 1960).
- The evidence is evolutionary, neurophysiological, and developmental.
- It raises the possibility that language is much more closely related to embodied cognition than current linguistic theories of grammar suggest.
- I'm going to argue that practically every aspect of language reflects this connection transparently, via object-oriented action concepts.
- The talk discusses this connection in terms of planning as it is viewed in Robotics and AI, with some attention to applicable machine learning techniques.
- Work In Progress under EU FP6 IP PACO-PLUS

Introduction

- The paper will define a path between representations at the level of the sensory manifold and perceptron learning to the mid-level of plans and explanation-based learning, and on up to the level of language grammar and parsing model learning.
- At the levels of planning and linguistic representation, two simple but very general combinatory rule types, **Composition** (the operator **B**) and **Type-Raising** (the operator **T**) will appear repeatedly.
- In planning terms,
  - composition **B** is seriation, while
  - type-raising **T** is object-oriented affordance

I: Planning and Affordance

- Apes really can solve the monkeys and bananas problem, using tools like old crates to gain altitude in order to reach objects out of reach.
- Such planning involves
  - Retrieving appropriate actions from memory (such as piling boxes on top of one another, and climbing on them),
  - Sequencing them in a way that has a reasonable chance of bringing about a desired state or goal (such as having the bananas).
  - Remembering good plans.

Figure 1: Ape and Bananas (Köhler 1925)

Figure 2: There is Another Approach (Köhler 1925)

Planning and Affordance

- Köhler showed that, in apes at least, such search seems to be
  - *reactive* to the presence of the tool, and
  - *forward-chaining*, working forward from the tool to the goal, rather than backward-chaining (working from goal to tool).
- The first observation implies that actions are accessed via perception of the objects that mediate them—in other words that actions are represented in memory *associatively*, as properties of objects—in Gibson's 1966 terms, as *affordances* of objects.
- The second observation suggests that in a cruel and nondeterministic world it is better to identify reasonably highly valued states that you have a reasonable chance of getting to than to optimize complete plans.
- Animal planning therefore involves *searching* through possible causally-related futures generated by the affordances of the available objects in the situation that obtains.

Planning and Affordance

- The problem of planning can therefore be viewed as the problem of finding a sequence of actions α, β, etc. in a “Kripke model”:

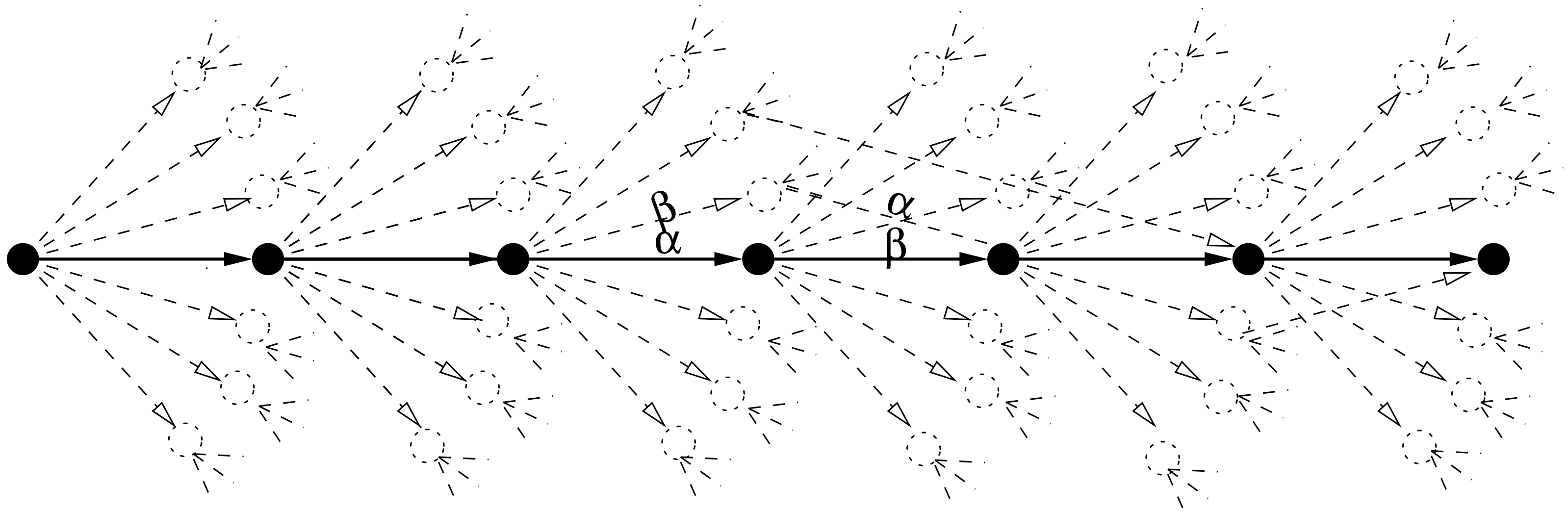

Figure 3: S4 Kripke Model of Causal Accessibility Relation

Of course, to plan safely in a cruel world, you also need a probabilistic model of success for α, β, and the ability to replan in real time when things go wrong.

Representing LDEC Operators

- We can think of actions as STRIPS operators or as finite-state transducers (FSTs) over (sparse) state-space vectors
- FSTs are closed under composition, and can be represented as simple neural computational devices such as Perceptrons, or the Associative Network or Willshaw Net (Willshaw 1981 cf. Marr 1969), which is specialized for representing associations between sparse vectors.
- The autoassociative network represents the association between situations, associating partial state vectors with the actions they afford.
- Similarly, the LDEC version of STRIPS update rule can be represented as a hetero-associative Willshaw net whose output specifies changes (Mourão *et al.*, cf. Modayil and Kuipers 2007; Amir and Chang 1968).

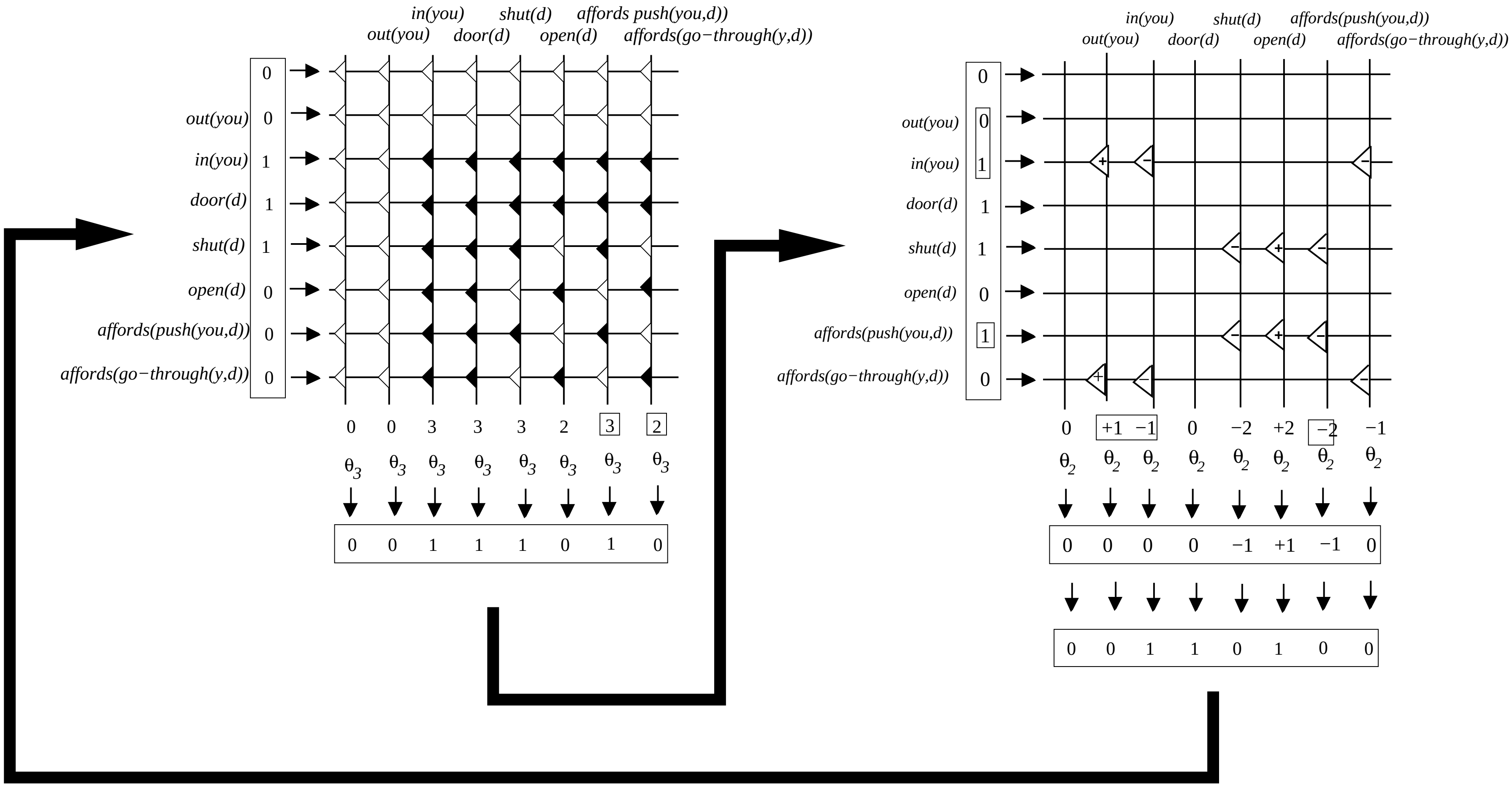

Figure 4: Planning cycle: Retrieving the affordance of *push* from the autoassociative net, generating the next state from the heteroassociative net, updating the state vector, and preparing to retrieve the affordance of *go-through*.

Learning with the Kernel Perceptron

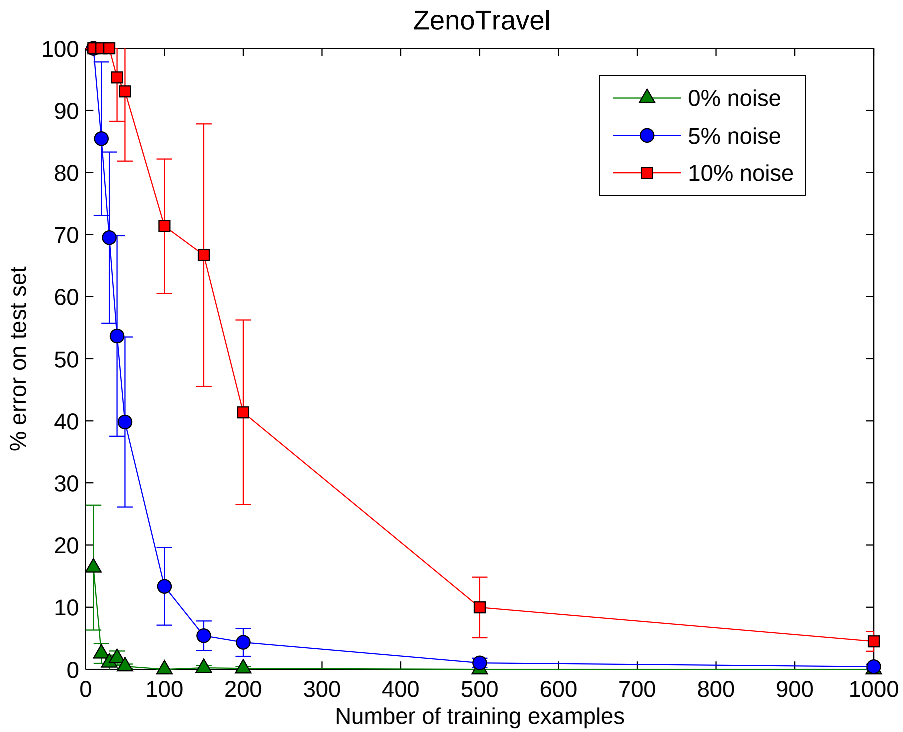

Figure 5: Learning results in noisy and fully observable versions of the ZenoTravel domain. Noise at level p% was simulated by flipping each bit in the state vector with probability p. The test sets were noiseless, fully observable sequences of observations and actions, of length 2000 (Mourão *et al.* 2009).

Reducing Complexity using Attention

- We need the Kernel Generalization of Perceptrons for this (Mourão *et al.* 2008, 2009).

- Complexity is $O(n^2)$, so we need to keep the state vector small.

- One "reactive" way to do this is to confine the elements to *fluents* and related *preconditions* of each action associated with perceptually evident objects.

- The "deictic" or attentional representations of Agre and Chapman (1987) and Pasula *et al.* (2007) are related.

II: From Planning to Language

- How do we get from seriation and affordance to language?
- An *action* (OAC) is a function from (partially specified) states to states
- An *affordance* is a (typed, curried) function from (typed) entities to events involving those entities.
- An *object* is a function from entities into type-raised functions **T***entity* from affordances into the results of applying them to the entity.
- Reactive planning is *seriation* or composition **B** of events of type *state* → *state* formed by applying objects to entities and affordances.

(1) $get\!-\!out \equiv \mathbf{B}(\mathbf{T}door_{19}\ go\!-\!through)(\mathbf{T}door_{19}\ push)$

Example

- The ape's concept of a box is a function from boxes into events *the box falling*, *their climbing-on the box* and *their putting the box on another box*, whose outcome the ape can evaluate, and forward-chain over.

- The affordances are of the following (Curried) types, where e is the type of an entity and t is the type of a state:[a]
  - $fall_{e \rightarrow t}$,
  - $climb\text{-}on_{e \rightarrow (e \rightarrow t)}$
  - $put\text{-}on_{e \rightarrow (e \rightarrow (e \rightarrow t))}$

- Thus the ape's box concept can be viewed as a set of object-concepts of type
  - $box1_{(e \rightarrow t) \rightarrow t}$
  - $box2_{(e \rightarrow (e \rightarrow t)) \rightarrow (e \rightarrow t)}$
  - $box3_{(e \rightarrow (e \rightarrow (e \rightarrow t))) \rightarrow (e \rightarrow (e \rightarrow t))}$

[a] Languages like Navajo with elaborate verb-classifier systems show we need a richer ontology!

Lexicalizing Affordance: Type Raising

- These functions are the result of Type-raising an object of type e over one-, two, and three-place functions.

- The mathematical concept of Type-Raising is closely related to the linguistic concept of grammatical relation or Case.

- For example, $box1_{(e \rightarrow t) \rightarrow t}$, a function over predicates like $fall_{e \rightarrow t}$ corresponds to a *subject*, marked in Latin or Japanese by Nominative Case.

Combinatory Categorial Grammar

- CCG eschews language-specific syntactic rules like (2) for English.

(2) $S \rightarrow NP\ VP$
$VP \rightarrow TV\ NP$
$TV \rightarrow \{proved,\ finds,\ \ldots\}$

- Instead, all language-specific syntactic information is *lexicalized*, via lexical entries like (3) for the English transitive verb:

(3) met $:= (S\backslash NP)/NP$

- This syntactic "category" identifies the transitive verb as a function, and specifies the type and directionality of its arguments and the type of its result.

Type Raising as Case

- We will assume that type-raising in the form of case is a universal primitive of grammar, as it is of planning.

All noun-phrases (NP) like "Harry" are polymorphically type-raised.

- In Japanese and Latin this is the job of case morphemes like nominative -*ga* and -*us*.
- In English NPs are ambiguous as to case, and must be disambiguated by the parsing model.

Syntactic Derivation

- (4)

$$
\begin{array}{ccc}
\text{Harry} & \text{met} & \text{Sally} \\
\overline{S/(S\backslash NP)}^{>\mathbf{T}} & \overline{(S\backslash NP)/NP} & \overline{(S\backslash NP)\backslash((S\backslash NP)/NP)}^{>\mathbf{T}} \\
 & \multicolumn{2}{c}{\overline{\qquad\qquad\qquad S\backslash NP \qquad\qquad\qquad}^{<}} \\
\multicolumn{3}{c}{\overline{\qquad\qquad\qquad\qquad\qquad S \qquad\qquad\qquad\qquad\qquad}^{>}}
\end{array}
$$

Relativization

- (5) that $:= (N\backslash N)/(S/NP)$

(6) (The woman) that Harry met

that	Harry	met
$(N\backslash N)/(S/NP)$	$S/(S\backslash NP)$ (>T)	$(S\backslash NP)/NP$
	S/NP (>B)	
$N\backslash N$ (>)		

(7) (The woman) that Harry says he met

that	Harry	says	he	met
$(N\backslash N)/(S/NP)$	$S/(S\backslash NP)$ (>T)	$(S\backslash NP)/S$	$S/(S\backslash NP)$ (>T)	$(S\backslash NP)/NP$
	S/S (>B)		S/S (>B)	
	S/NP (>B)			
$N\backslash N$ (>)				

CCG reduces the linguists' MOVE to adjacent MERGE

Coordination

- (8) give Harry a book and Sally a record

$$
\begin{array}{cccccc}
\text{give} & \text{Harry} & \text{a book} & \text{and} & \text{Sally} & \text{a record} \\
\overline{DTV} & \overline{TV\backslash DTV}^{<\mathbf{T}} & \overline{VP\backslash TV}^{<\mathbf{T}} & \overline{(X\backslash X)/X} & \overline{TV\backslash DTV}^{<\mathbf{T}} & \overline{VP\backslash TV}^{<\mathbf{T}} \\
 & \multicolumn{2}{c}{\overline{VP\backslash DTV}^{<\mathbf{B}}} & & \multicolumn{2}{c}{\overline{VP\backslash DTV}^{<\mathbf{B}}} \\
 & & & \multicolumn{3}{c}{\overline{\qquad\qquad}^{>}} \\
 & \multicolumn{5}{c}{\overline{VP\backslash DTV}^{<}} \\
\multicolumn{6}{c}{\overline{VP}^{<}}
\end{array}
$$

CCG reduces the linguists' COPY/DELETE to adjacent MERGE

Syntax = Type-Raising and Composition

- CCGs combination of type-raising and composition yields a "nearly context-free" permuting and rebracketing calculus closely tuned to the needs of natural grammar.

- The argument cluster coordination construction (8) is an example of a universal tendency for "deletion under coordination" to respect basic word order: in all languages, if arguments are on the left of the verb then argument clusters coordinate on the left, if arguments are to the right of the verb then argument clusters coordinate to the right of the verb (Ross 1970):

(9)	SVO:	*SO and SVO	SVO and SO
	VSO:	*SO and VSO	VSO and SO
	SOV:	SO and SOV	*SOV and SO

Conclusion

- The lexicon is the only locus of language specific infomation in the theory of grammar.
- The universal projective syntactic component of natural language grammar is based on the combinators **B**, **T**.
- These combinators are provided ready-made, by a sensory motor planning mechanism that we share with a number of animals.
- OACs are a nice way to think about planning

Appendix

- If apes have **B**, **T**, why don't chimpanzees have human linguistic capability?
  - In particular, why do they appear to lack a truly recursive syntax?
  - Since the apes have everything necessary for syntactic projection (not to mention mirror neurons and FOXP2), the only possible locus for the difference is the human lexicon.
  - Specifically, some distinctively recursive concepts that humans lexicalize there.
  - The main contender is the human concept of other minds (Tomasello 2001).
- If so, the origin of recursion in syntax is essentially semantic.

References

Agre, Phillip and Chapman, David, 1987. "Pengi: An Implementation of a Theory of Activity." In *Proceedings of the Sixth National Conference on Artificial Intelligence (AAAI-87)*. Los Altos, CA: Morgan Kaufmann.

Amir, Eyal and Chang, Allen, 1968. "A Memory Storage Model Utilizing Spatial Correlation Functions." *Kybernetik* 5:113–119.

Gibson, James, 1966. *The Senses Considered as Perceptual Systems*. Boston, MA: Houghton-Mifflin Co.

Köhler, Wolfgang, 1925. *The Mentality of Apes*. New York: Harcourt Brace and World.

Lashley, Karl, 1951. "The Problem of Serial Order in Behavior." In L.A. Jeffress (ed.), *Cerebral Mechanisms in Behavior*, New York: Wiley. 112–136. reprinted in Saporta (1961).

Marr, David, 1969. "A Theory of Cerebellar Cortex." *Journal of Physiology* 202:437–470. Reprinted in Vaina 1991.

Miller, George, Galanter, Eugene, and Pribram, Karl, 1960. *Plans and the Structure of Behavior*. New York, NY: Henry Holt.

Modayil, Joseph and Kuipers, Ben, 2007. "Autonomous development of a grounded object ontology by a learning robot." In *Proceedings of the AAAI Spring Symposium on Control Mechanisms for Spatial Knowledge Processing in Cognitive/Intelligent Systems*. AAAI.

Mourão, Kira, Petrick, Ron, and Steedman, Mark, 2008. "Using Kernel Perceptrons to Learn Action Effects for Planning." In *Proceedings of 3rd International Conference on Cognitive Systems (CogSys 2008)*. University of Karlsruhe, 45–50.

Mourão, Kira, Petrick, Ron, and Steedman, Mark, 2009. "Learning Action Effects in Partially Observable Domains." In *Proceedings of the ICAPS 2009 Workshop on Planning and Learning*. Thessaloniki, Greece, 15–22.

Pasula, Hanna, Zettlemoyer, Luke, and Kaelbling, Leslie, 2007. "Learning Symbolic Models of Stochastic Domains." *Journal of AI Research* 29:309–352.

Piaget, Jean, 1936. *La naissance de l'intelligence chez l'enfant*. Paris: Delachaux et Niestle. translated 1953 as *The Origin of Intelligence in the Child*, Routledge and Kegan Paul.

Ross, John Robert, 1970. "Gapping and the Order of Constituents." In Manfred Bierwisch and Karl Heidolph (eds.), *Progress in Linguistics*, The Hague: Mouton. 249–259.

Saporta, Sol (ed.), 1961. *Psycholinguistics: A Book of Readings*. New York: Holt Rinehart Winston.

Steedman, Mark, 2000. *The Syntactic Process*. Cambridge, MA: MIT Press.

Tomasello, Michael, 2001. "Perceiving Intentions and Learning Words in the Second Year of Life." In Melissa Bowerman and Stephen Levinson (eds.), *Language Acquisition and Conceptual Development*, Cambridge: Cambridge University Press. 132–158.

Vaina, Lucia (ed.), 1991. *From Retina to Neocortex: Selected Papers of David Marr*. Boston, MA: Birkhauser.

Willshaw, David, 1981. “Holography, Association and Induction.” In Geoffrey Hinton and James Anderson (eds.), *Parallel Models of Associative Memory*, Hillsdale, NJ: Erlbaum. 83–104.